WORKED EXAMPLE

Let's say that you want to create Table B08406, "Sex of workers by means of transportation to work for workplace geography," for the state of Alaska from the files on the ftp site http://www2.census.gov/acs2008_1yr/summaryfile/. Which files do you need? How do you read the files?

You will need three files:
1. The data dictionary (merge_5_6.sas7bdat)
2. The data file (e2008ak0001000.txt)
3. The geography file (g20081ak.txt)

Start with the data dictionary, merge_5_6.sas7bdat. Under the "Tblid" column, look for the value "B08406". You will see that the "Sequence Number" is "0003." This means that the data you looking for are in the data file "e2008ak0003000.txt". How do you know this is the right file? You know this from the name of the file: the "e" stands for estimate, 2008 is the year, "ak" is the state (Alaska), and "0001" is the sequence number (which contains the data for Table B08406). See the "File Naming Conventions" section in Chapter 2.

Then use the geography file for Alaska to determine the location within the state to which the data refer. The appropriate file is g20061ak.txt, where "g" means "geography", 2008 is the year, 1 is the period estimate (in this case, 1-year estimate), and ak is the state. (For each state, the geography file contains the lower-case postal-abbreviation of the state.)

When you open the data file, e20081ak0001000.txt, you will see the following comma-delimited fields on the first line:

ACSSF,2008e1,ak,000,0003,0000001,335039,264535,219908,44627,35706,…

The first six fields – from "ACSSF" to "0000001" – are identifiers.
A. The first field tells you that this is an ACS Summary File;
B. The second tells you that these data are one-year estimates for the year 2008 (notice the "e" before "2008" and the "1" at the end);
C. The third tells you the state ("ak" is Alaska);
D. The fourth is an iteration number;
E. The fifth is the sequence number,
F. The last is a logical record code (LOGRECNO). The LOGRECNO identifies the location within a state.

The geography file, g20081ak.txt, defines the LOGRECNO. Each LOGRECNO specifies a location pertaining to the state. For example, a LOGRECNO of "0000001" means the state of Alaska; a LOGRECNO of "0000002" means just the urban areas in Alaska; a LOGRECNO of "0000003" refers to just rural areas in Alaska. Notice that each

state has its own geography file.

The other fields in the data file, from the seventh on, are data values. Each field corresponds to the value of the "line number" variable in the data dictionary. So field number seven (the 335039 value, after the sixth comma) corresponds to line number one, which is "Total". Field number eight (the 264535 value, after the seventh comma) refers to line number two, which is "Car, Truck, or Van." Field number nine (the 219908 value) corresponds to line number three, which is "Drove alone." This continues all the way up to line number 51, at which point Table B08406 ends.

Were you to read this into a computer program using software such as SAS, you could translate the first nine fields of e20081ak0001000.txt as follows:

TABLE 8406:
SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK FOR WORKPLACE GEOGRAPHY

Type of File	Estimate Type	State	Iteration Number	Sequence Number	Logical Record Number	Total	Car, Truck, or Van	Drove Alone
ACSSF	2008e1	ak	000	0003	0000001	335039	264535	219908
ACSSF	2008e1	ak	000	0003	0000010	149452	134141	113696
ACSSF	2008e1	ak	000	0003	0000011	47314	39757	33343
ACSSF	2008e1	ak	000	0003	0000012	.	.	.
ACSSF	2008e1	ak	000	0003	0000013	149452	134141	113696
ACSSF	2008e1	ak	000	0003	0000017	149452	134141	113696
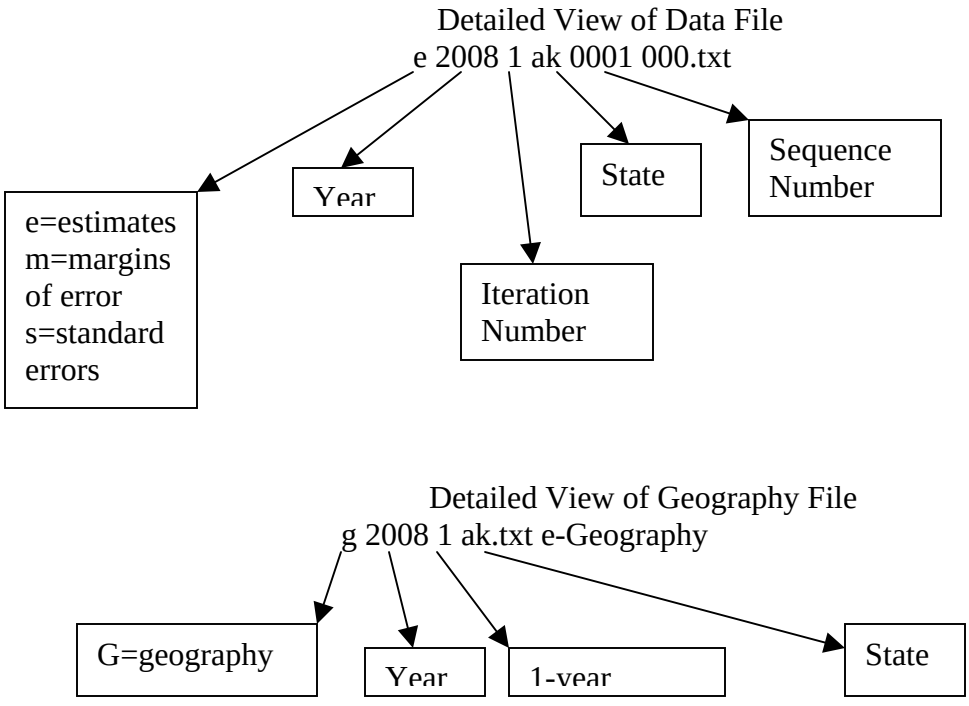
Figure 2.4 Linking Geographic Header File to the Data Files
Worked Example: User wants to get all of the data for Table B08406, which is found on the file with SEQ=0003. Show the naming convention for all files.